Journal of Biomedical Informatics 56 (2015) 292–299

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Using natural language processing to provide personalized learning opportunities from trainee clinical notes

Joshua C. Denny [a,b,*], Anderson Spickard III [a,b], Peter J. Speltz [a], Renee Porier [c,f], Donna E. Rosenstiel [c], James S. Powers [b,c,d,e]

[a] *Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, United States*
[b] *Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, United States*
[c] *The Center for Quality Aging, Vanderbilt University School of Medicine, Nashville, TN, United States*
[d] *The Meharry Consortium Geriatric Education Center, Meharry Medical Center, Nashville, TN, United States*
[e] *The Tennessee Valley Geriatric Research Education and Clinical Center, Tennessee Valley Healthcare System, Nashville, TN, United States*
[f] *Office of Health Sciences Education, Vanderbilt University School of Medicine, Nashville, TN, United States*

ARTICLE INFO

Article history:
Received 13 August 2014
Revised 1 June 2015
Accepted 3 June 2015
Available online 10 June 2015

Keywords:
Medical education
Natural language processing
Decision support
Geriatric education
Altered mental status
Advanced directives

ABSTRACT

Objective: Assessment of medical trainee learning through pre-defined competencies is now commonplace in schools of medicine. We describe a novel electronic advisor system using natural language processing (NLP) to identify two geriatric medicine competencies from medical student clinical notes in the electronic medical record: advance directives (AD) and altered mental status (AMS).

Materials and methods: Clinical notes from third year medical students were processed using a general-purpose NLP system to identify biomedical concepts and their section context. The system analyzed these notes for relevance to AD or AMS and generated custom email alerts to students with embedded supplemental learning material customized to their notes. Recall and precision of the two advisors were evaluated by physician review. Students were given pre and post multiple choice question tests broadly covering geriatrics.

Results: Of 102 students approached, 66 students consented and enrolled. The system sent 393 email alerts to 54 students (82%), including 270 for AD and 123 for AMS. Precision was 100% for AD and 93% for AMS. Recall was 69% for AD and 100% for AMS. Students mentioned ADs for 43 patients, with all mentions occurring after first having received an AD reminder. Students accessed educational links 34 times from the 393 email alerts. There was no difference in pre (mean 62%) and post (mean 60%) test scores.

Conclusions: The system effectively identified two educational opportunities using NLP applied to clinical notes and demonstrated a small change in student behavior. Use of electronic advisors such as these may provide a scalable model to assess specific competency elements and deliver educational opportunities.

1. Introduction

Attaining clearly defined competencies is essential for health care professionals. National and international accreditation bodies expect training programs to document the attainment of these competencies by learners [1–4], and with the growing population of elders, more attention has been paid recently to geriatric competency [5,6]. To measure competency-specific performance of medical trainees, educators have focused primarily on standardized milestones and exams, observed structured clinical skills exams, mostly manual logs of clinical exposures, and clinical assessment by attending physicians. Electronic medical records (EMRs) may enable an automated approach that captures students' clinical experiences as a byproduct of their normal clinical work. As a test of a new paradigm for delivering medical education content, we developed an automated education advisor system that analyzed students' EMR notes for relevance to two geriatric competencies and then emailed customized feedback. The two competencies, part of 26 geriatric competencies identified by the American Association of Medical Colleges (AAMC) [5], included: assessment of advanced directives (AD) and evaluation of patients with altered mental status (AMS). We evaluated the accuracy of the system and the effect on each student's knowledge through multiple choice tests.

* Corresponding author at: 2525 West End Ave, #6049, Nashville, TN 37203, United States. Tel.: +1 615 936 5034.
E-mail address: josh.denny@vanderbilt.edu (J.C. Denny).

http://dx.doi.org/10.1016/j.jbi.2015.06.004
1532-0464/

2. Background

2.1. Geriatric competencies for older adults

The AAMC and the John A. Hartford Foundation (JAHF) developed a minimum set of graduating medical student competencies to assure competent care for older patients [5]. Similar efforts have generated competencies for residents and nursing programs (http://www.pogoe.org/geriatrics-competencies). Using the input of leading geriatric educators and survey responses from educators in a number of clinical domains, the consensus panel established eight core geriatric competency domains. Each competency domain contains 2–5 competencies, outlining detailed milestones for medical students in each domain. These competency domains represent an agreed-upon framework to guide curriculum and assessment of medical students. For example, the "cognitive and behavioral disorders" domain includes 5 competencies, which deal with clinical presentation, differential diagnosis, evaluation, and treatment of delirium, dementia, and depression in older adults. These are common diagnoses contributing to AMS presentations in the older adult. Similarly, it is a goal that graduating medical students understand ADs. All hospitalized patients should be assessed for ADs as a result of the Patient Self Determination Act [7]. Given the heightened importance of ADs for elders, we set a clerkship goal that students should discuss ADs with all patients 65 years and older. Competencies for AMS and AD were the target of the present study.

2.2. Competency assessment

Competency-based assessment methods often combine a variety of modalities to provide a comprehensive evaluation of a learner's knowledge and proficiency [8,9]. Medical schools using competency-based assessments typically rely on education portfolios to track each student's progress [10–14]. Portfolio components can include personal reflections on experiences, examinations and their scores, individual and small group projects, simulation encounter reports such as observed structured clinical examinations (OSCEs), mentoring experiences, and clinical exposures. Handwritten log books [15,16] or electronic logs [17,18] allow students to record patient information including demographics, diagnosis, procedures performed, and/or severity of illness. Capture rates of learners' experiences are low because trainees are too busy to enter the data into the system [15,19]. Further, teachers often disagree with students on the primary diagnosis of the case [16,20].

2.3. Overview of the learning portfolio system

To address some of the concerns with manual logs and provide a more robust capture of a trainee's clinical exposure, we developed the KnowledgeMap Learning Portfolio ("Portfolio") system at Vanderbilt, which collects all trainee-authored clinical notes from the EMR as well as providing a forum for other portfolio activities such as personal reflections and specific course content [21]. Medical students are required to write notes (e.g., history and physicals, progress notes, discharge summaries) in the EMR on their patients during their clinical years. From this record, Portfolio automatically creates procedure logs and catalogs patient exposures. Mentors (typically attending and resident physicians) can provide feedback on clinical notes through Portfolio, which has been shown to increase frequency of feedback and improve the quality of the student's assessment and plan in their notes [21]. By applying natural language processing (NLP) through use of the KnowledgeMap concept indexer [22] to identify Unified Medical Language System (UMLS) concepts [23] and SecTag [24,25] to identify note section headers, we have developed search algorithms to automatically map clinical notes to school-identified learning objectives [14,26]. In use since 2005, Portfolio currently includes nearly 5 million indexed trainee-authored notes.

3. Materials and methods

3.1. Setting

We approached third-year medical students at Vanderbilt University School of Medicine during their medicine clerkships between January 2010 through December 2010 with the opportunity to receive learning opportunity messages related to geriatric competencies based on their EMR clinical notes. All students spend at least half of the medicine experience at Vanderbilt University Hospital, the setting for this study; the rest of the patient exposures occur at the Nashville Veterans Affairs hospital whose records were not available to Portfolio. All students in these clerkships were offered the chance to enroll. Students who chose to participate in the study completed a 22-item multiple choice test of geriatric clinical knowledge at the beginning and end of the clerkship. Students enrolled in the study received email alerts for clinical notes matching either of the competencies after completion of the first exam until the completion of their medicine rotation. The Vanderbilt Institutional Review Board approved this study and all participants completed informed consent prior to participation. Course instructors were blinded as to student participation status.

3.2. Education advisor system

We developed a generic framework to allow for creation of electronic education advisors, and then developed specific rules for two geriatric education advisors: altered mental status and advanced directives. These rules were stored in a database and included: a list of students to evaluate; minimum and maximum patient ages to consider; a list of concepts (as UMLS concept unique identifiers) and words to search for within a note; note sections to interrogate; a minimum score threshold required to consider the note a match to the advisor (based on the numbers of concepts and words identified via NLP); boilerplate text for the advisor consisting of "key facts" to send in each email; and whether or not to include a list of relevant medical school curriculum documents relevant to the student's note. The advisor system is flexible, driven by configurable options stored in a database. Search queries and document collections can be built through the KnowledgeMap and Portfolio web interfaces.

Once a student writes a note in the EMR, it was immediately sent from the EMR to Portfolio (as done for all trainee notes as part of the Portfolio system design), which indexed it for all UMLS concepts, tagged with note section according to the SecTag hierarchy of section headers [25]. Information is stored in a relational database. For each advisor, Portfolio analyzed nightly all new trainee notes matching patient characteristics and containing the specified UMLS concepts (see Appendix) located in the specified sections of notes from among all identified UMLS concepts stored in Portfolio (e.g., the core concepts indexed from each note were not restricted to those deemed relevant to the advisors in order to maintain maximum flexibility for other use cases, such as future advisors and searching). Notes exceeding the score threshold resulted in emails to the student with content generated based on the settings, including a link to the relevant note (without personal health identifiers). If the score threshold were set to 0, any note matching the patient characteristics would result in an email, and different text was generated based on whether concepts were or were not present (such as with the AD advisor, see below). All generated emails were logged. A maximum of one of each alert type could be sent to each student per patient.

If the rule specifies that relevant document should be sent, a list of the top 10 most relevant documents would be sent by querying all documents listed in the KnowledgeMap curriculum management system, a web-based system in which all curriculum content (e.g., slide shows and lecture handouts) is mapped to UMLS concepts through the same NLP system [27]. For this study, this functionality was used only for the AMS advisor. First, to attempt to focus on documents addressing AMS topics relevant to the case, the system identified UMLS concepts found in the clinical note that overlapped the AMS concepts defined in the advisor (see Appendix for an example of recommended documents for an AMS alert). For example, the UMLS indices of curricular documents contain all matching UMLS concepts, most of which are irrelevant to AMS. By ranking curricular documents only by their overlap to AMS concepts found in the note, we attempted to identify curricular documents specifically addressing AMS management, workup, and differential diagnosis found in the case (e.g., malignancy, infection, hyponatremia). A list of this broader set of 755 AMS related concepts covering AMS-causing diagnoses, treatments, signs and symptoms, and workup are found in Supplementary Table 3. These overlapping concepts were used as the search query to identify curriculum documents from the entire set of ~25,000 KM curricular documents. Curriculum documents were ranked using the term frequency, inverse document frequency (TF–IDF) method, and the top 10 were included in the email sent to the student. In this way, the system attempts to select documents that addressed most specifically the relevant AMS topics in the note, not other aspects of the case.

3.3. AMS and AD advisors

Table 1 summarizes the basic characteristics of both advisors. Examples of both advisors are provided in the Appendix. The AMS advisor looked for 83 UMLS concepts and two keywords (“AMS”, “altered mental status”) in the chief complaint, history of present illness, past medical history, physical exam, and assessment and plan sections of the notes (including relevant subsections); other sections were ignored. The score threshold was set to 1, meaning that notes with at least one matching concept in these sections were considered to be relevant to AMS, and in those cases, an email would be sent to the student consisting of boilerplate text and relevant curriculum documents and document collections. The AMS advisor included text specifying which concepts in the student’s note triggered the AMS advisor, and then links to other content were customized based on the approaches listed above.

The AD advisor analyzed the first student note for each patient they encountered who was ⩾65 years old. The subject line and content of the advisor varied based on the presence or absence of AD concepts in the student’s note. For notes in which AD concepts were found, the system generated emails with the subject “Portfolio feedback: Advanced Directives” and note content included: “This note appears to contain a discussion of Advanced Directives.” In notes in which AD concepts were not found, the subject was “Portfolio feedback: Advanced Directives not discussed in your note” and note content included: “This note does not appear to contain a discussion of Advanced Directives.” Both emails included key facts about advanced directives and links to relevant documents in the curriculum (which, unlike the AMS advisor, were the same ten documents regardless of the content of the student’s note).

All emails and notes evaluated were logged such that students were never sent the same advisor on the same patient twice.

3.4. Evaluation

We analyzed the system in two ways: advisor accuracy and educational impact. Advisor accuracy to correctly identify a target note was judged via recall and precision. All AMS and AD advisors were evaluated by two board-certified physicians (AS, JP) to determine precision. Since AD advisors fired for all patients ⩾65 years old, both recall and precision could be calculated by reviewing the student notes for all patients ⩾65 years old to determine which contained discussions of AD. The physicians reviewed student notes blinded to the determination of the algorithm. For the AMS advisor, the physicians reviewed all AMS advisors for correctness and then a random sample of 108 notes were reviewed to estimate recall (a comprehensive review of all student notes generated in the clerkship would have resulted in review of >7000 notes since

Table 1
Characteristics of educational advisors. Characteristics marked by asterisks (*) are configurable options in the database.

	Advanced Directives (AD)	Altered Mental Status (AMS)
Purpose	To educate students about AD; remind students to write advance directives for patients ⩾65 years old	To provide “just in time” education content on patient with AMS; to educate on the differential diagnosis and treatment of AMS
Patient age requirement (years)*	⩾65	⩾18
Email subject line {optional text}*	Portfolio feedback: Advanced Directives {not discussed in your note}	Portfolio feedback: Altered Mental Status
Contents in body of email*	• Informs patient whether AD concepts were found in their note or not • Brief summary about performing ADs and what constitutes an AD	• AMS concept(s) found in their note • Brief summary of AMS presentation and differential diagnosis
Prespecified documents sent*	8 prespecified documents related to AD	18 prespecified AMS documents
Relevant documents sent*	N/A	10 curricular documents most related to the patient’s AMS concepts identified in the note
Other content*	N/A	List of prespecified AMS searches for relevant causes, ranked by relevance to patient’s presentation (e.g., metabolic diseases, substance use, Parkinson’s)
Example triggering words (in quotes)/UMLS concepts	Hospice, living will, durable, advanced directive, do not resuscitate/do not intubate orders (see Supplemental Table 1 for full list)	“AMS”, “altered mental status” and 83 concepts, e.g.: Wernicke’s encephalopathy, hallucinations, dementia, delirium, altered mental status, delusions, Hepatic Encephalopathy (see Supplemental Table 2 for full list)
Example concepts for ranking documents	N/A	AMS search concepts, plus 755 additional concepts, e.g.: hypercalcemia, malignancy, hyponatremia, sepsis, infections, etc. (see Supplemental Table 3 for full list)

AMS could develop at anytime during a hospitalization). In the random sample, physicians were blinded to the determination of the algorithm. Reviewers judged precision by reviewing the corresponding clinical note to determine if the note truly contained a reference to AMS (e.g., including both active problems during the admission or chronic treatment problems, such as a patient with treated dementia even if not a primary reason for admission).

To determine the accuracy of the documents suggested by the system, the reviewers (AS, JP) then reviewed all suggested curriculum documents for 78 randomly-chosen notes for which 769 note-document pair suggestions were made. Twenty note-document pair suggestions were overlapped between the two reviewers to enable calculation of interrater agreement. Review was done via a website interface that allow rating of relevant and not relevant. Suggested documents were judged not relevant if they addressed a topic not relevant to the AMS presentation in the case.

Two 22-item multiple choice question examinations (MCQ) for pre and post testing were derived from published geriatric exams [28–30], mapped to the AAMC geriatric competency domains. The tests were beta tested on fourth year medical students during the preceding third year Medicine Clerkship (2009) and contained six questions that were identical but located at different positions within the test. The pre-test was administered at the start of the medical clerkship and post-test at the end of the clerkship.

Following completion of the study, we used the AD algorithm to identify subsequent mentions of ADs after students first received an email alert in which they had not mentioned an AD. Emails were not issued at this time, but notes were evaluated for presence and absence of AD concepts.

4. Results

Fig. 1 presents the overview of the students approached and who enrolled. Sixty-six students agreed to be in the study. Students wrote a total of 7124 notes on 2359 adult patients during their medicine clerkship; 765 (32%) of these patients were over 65 years old. About 40% of these patients were seen while the students were enrolled in the study. Overall, 393 emails were sent to 54 students (82% of those enrolled), including 270 for AD and 123 for AMS (Table 2). Fig. 2 shows the workflow and an example of one of the advisors that fired for a student–patient combination. Out of 270 notes considered by the algorithm on patients $\geqslant$65 years old, only 35 (13%) had mentions of AD topics. The algorithm correctly detected 24 mentions of AD topics (true positives) in notes without any false positives; however, 11 mentions of AD topics were not identified by the system (false negatives). All 235 notes without AD mentions were correctly detected as such by the system (true negatives). The most common cause of false negatives resulted from NLP errors (e.g., "will" being interpreted as a verb instead of a noun). Thus, the recall and precision for the AD algorithm were 67% and 100%, respectively. Notably, all mentions of ADs occurred in student notes written after they received an AD advisor triggered because they had not mentioned ADs in their initial note.

The AMS advisor sent 123 emails to 40 students. One hundred fourteen of these alerts were judged to correctly reflect AMS concepts in the note, yielding a precision of 93%. Errors were due to failed detection of negation signals, incorrect acronym disambiguation (e.g., "NPH" misinterpreted as "normal pressure hydrocephalus" instead of a type of insulin), and discussions of AMS-related concepts in assessment sections that were not conditions experienced by the patient (e.g., a hypothetical outcome, such as a medical student discussing the future possibility of encephalopathy in end-stage cirrhosis). Recall was assessed by blinded review of 108 random notes. Of these notes, 22 notes were judged relevant to AMS on 13 distinct patient–student pairs (the AMS advisor will only alert each student once for each patient, even if subsequent notes match the algorithm). The AMS advisor identified all 13 patients as being relevant to AMS, for an estimated recall of 100%.

The system selected and emailed links to 260 unique documents from the medical school curriculum in the 393 advisor emails sent to students. The same eight documents were included in every AD advisor (by design). The AMS advisor, which dynamically selected the 10 best-matched curricular documents for each student note in addition to a preselected list of 18 documents, suggested 234 unique documents. Students accessed educational links 34 times from the 393 email alerts; nine of these views were from the AMS advisor documents tailored to their note.

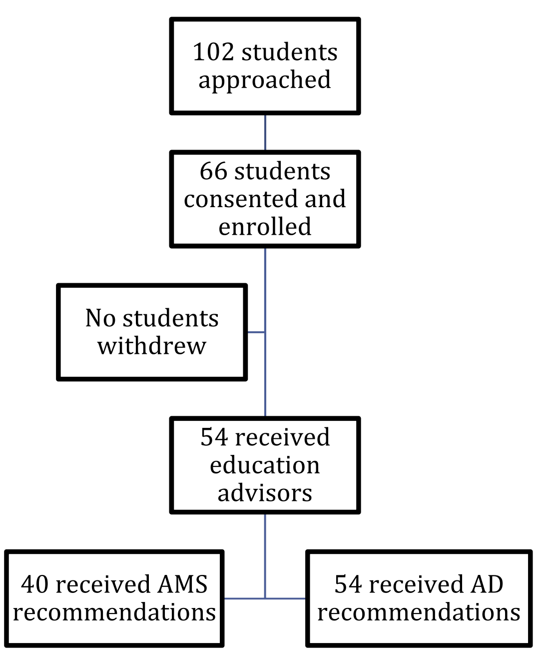

Fig. 1. Study enrollment.

Table 2
Evaluation of education advisors.

	Advanced directives	Altered mental status
Patient age (median, interquartile range)	74.4 (68–79)	59.7 (52–71)
Total emails sent	270	123
True positives[a]	24	114
False positives[a]	0	9
False negatives[a]	11	0 (estimate[b])
True negatives[a]	235	N/A
Unique patients triggering emails	261	121
Unique students receiving emails	54	40
Recall	69%	100% (estimate[b])
Precision	100%	93%

[a] True positives refer to accurate detection of AD or AMS mentions in notes. False negatives were notes that should have been identified as having AD mentions or related to AMS. For all patients 65 and older, AD emails were sent to patients (thus there are true negatives); however, no attempt to define true negatives for AMS was performed.

[b] This value is estimated from a review of 108 randomly selected notes during the review period; of these, 22 notes on 13 patients were found relevant to AMS and all were identified by the AMS advisor.

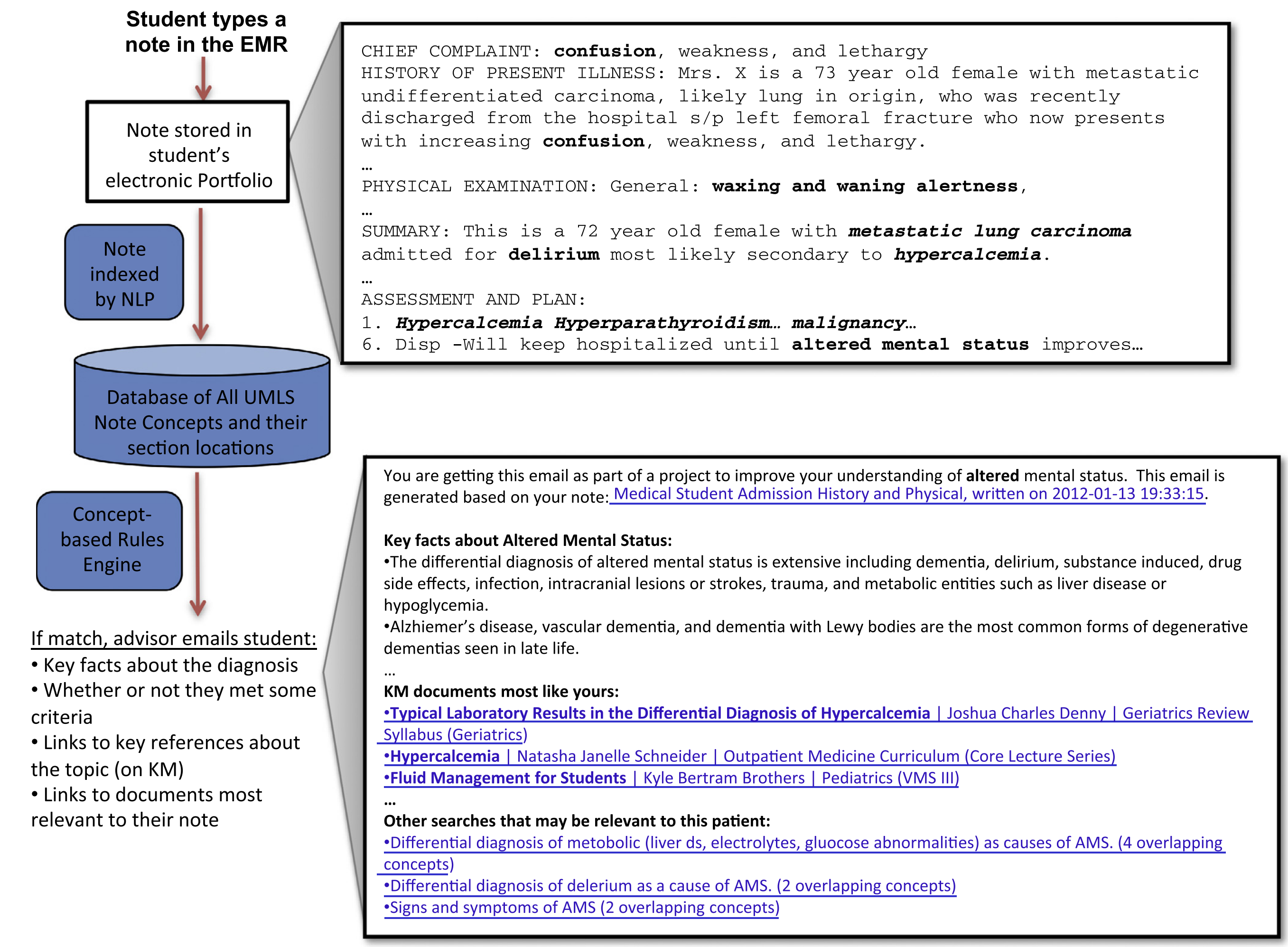

Fig. 2. Overview of automated education advisors system. Students type notes in the electronic medical record, and the note is then sent to the Learning Portfolio system, where they are concept indexed using an natural language processing system. A system of rules determines if the note meets criteria to alert the student, and then it sends an email to the student (only alerting that student once per rule per patient). The note is a shortened, de-identified example of a real note, and the email a shortened version of a real email sent for that note (with changed date). The **bold** text in the note represent concepts alerting the system to an altered mental status note, and the ***bold italicized*** text are concepts deemed relevant to the cause of altered mental status (used to determine which KnowledgeMap documents are most relevant). KM = KnowledgeMap.

Of the 226 patients in whose initial notes students did not mention AD, 48 students wrote subsequent notes on the same patient for a total of 148 patients. Analysis of these notes following the AD advisor email for a given patient revealed that eight (17%) students later mentioned AD concepts for eight of those patients. Interestingly, 6 of the 8 of the students who referenced AD concepts followed links in advisor emails at some point during the intervention.

Of the 769 suggested documents, 537 (69.8%) were judged relevant to the AMS diagnoses being discussed, evaluated, or treated in the note. The most common suggestions judged irrelevant were for documents suggested based on seizure concepts or for different populations. For example, a document on neonatal seizures was judged irrelevant to an older patient, and documents focusing on idiopathic epilepsy were not judged relevant to notes dealing with patients with medication or trauma-induced seizures. Additionally, seizures were often mentioned as a consideration in the differential but not considered strongly, and these were judged by reviewers as irrelevant. Similar errors occurred when other diagnoses were considered only vaguely as a list of possible or ruled out diagnoses. Finally, rarely documents were suggested relevant to the patient but not to AMS (e.g., hypertension management in a patient without hypertension as the cause of their AMS). The percent agreement in the 20 randomly-chosen note-document pairs was 100% between the two reviewers.

There was no difference in scores on the multiple-choice exams before (mean 62%) or after (mean 60%) the intervention. Restricting to questions regarding AD, the same two questions were asked before and after the clerkship, with a trend toward improvement in scores from 76% to 84% ($p = 0.39$). Different questions were asked for AMS between the pre and post tests. There was a small decrease in score, from 60% to 53% ($p = 0.05$), when restricting to AMS questions.

5. Discussion

We developed natural language processing algorithms to evaluate student notes and provide near real-time electronic feedback for two geriatric competencies. We found that the system performed well overall, with both systems achieving >90% precision, which we considered more important than recall in the design of this system. Overall, we observed that students rarely discussed ADs in patients on their initial interactions with older adults, mentioning them in only 13% of their notes. However, no student mentioned ADs without first having received an advisor email reminding them, and 17% of students included ADs in their subsequent notes on these patients not included in the evaluation. Thus, in this manner, the system had the desired goal of changing behavior, albeit a small effect. An advantage of this assessment is that it is measuring not knowledge but real clinical change in documentation. Given the low cost and burden of using such a system, we conclude that such education advisors may be a useful adjunct to other forms of instruction.

The tools used in this project could be applied to other sets of competencies across multiple modalities of electronic written work such as patient notes, curriculum materials, or reflection pieces along the continuum of medical training, although work to select target concepts and search queries do require some work. The search concepts, target note, sections, and score thresholds for this study were created and refined in a period of a few weeks involving several meetings of a team of individuals with clinical, educational and informatics expertise. Previous Portfolio studies have shown that such systems to facilitate student–mentor interactions result in better feedback to trainees and better assessment and plans in students' notes [21]. This work shows the promise of automatic capture and cataloguing of learners' experiences as well as proficiency assessments from a log of clinical notes. Similar approaches could be used for other competencies and with different EMR systems.

Students accessed linked educational content from 9% of the alerts sent. Others have noted that users' viewing of information resources is low in clinical information systems [31,32]. We anticipated low click-through rates, and thus embedded key facts about AD and AMS into the emails. Despite the low access rate, it is interesting to note that 6 of the 8 students who wrote ADs did access link content at some point, suggesting perhaps a higher level of interaction with the system. However, we are unable to evaluate whether students read the emails or not. In informal focus groups following the study, students told us that after the first few emails, the AD advisor subject line itself (reminding them whether or not they had written an AD or not) was sufficient for them. (The full text of an AMS and AD advisor are included in the Appendix.) In addition, overall geriatric medical knowledge as tested on the multiple-choice geriatric exam was not different. Possible explanations includes: (1) the exam broadly covered geriatric medicine, not just the narrow topics of the advisors, (2) the majority of the questions on the two forms of the exam differed, and (3) many personal and experiential factors can impact one's performance on a multiple choice test. On the AD portion of the test, which asked the same questions on both the pre- and post-test, there was a trend toward improvement. It is difficult to analyze the small decrease in performance for the AMS questions given that they were different. More careful assessments of knowledge changes are needed to discern specific improvement.

As expected, several of the errors for both algorithms resulted from NLP errors. One complication was that Portfolio's database of concepts does not currently contain the negation status of identified concepts (e.g., "no AMS"). Although the KnowledgeMap NLP system does consider negation currently [33], it did not when Portfolio was originally designed and prior Portfolio applications have not required negation status. This study indicates that the system should now be extended to record negation status for concepts. A few errors in recall for AD and AMS notes resulted from notes that failed to index properly by Portfolio due to irregular note types or characters embedded in the note.

Nearly 70% of the suggested curriculum documents were judged relevant to the AMS diagnoses being considered, evaluated, or treated in the note. It is important to note that our assessment was rather strict: for example, a document on hypertension management for a patient with hypertension but in whom hypertension was judged irrelevant to the cause of AMS was judged irrelevant. Some of the challenges encountered would be difficult to address by our simple text-matching scheme and would likely require other metadata elements in the search, such as filtering out pediatric documents for older patients. Another practical issue not addressed in this evaluation was that some of the document titles (e.g., "preceptor guide"; these were often very relevant documents) were not particularly helpful to a student when knowing whether or not to view a document. Encouraging faculty to create meaningful titles for curricular documents is an important practical result of this study to facilitate document searching by students.

We are unaware of other attempts to directly provide automated note feedback to trainees via NLP; however, there is a long history of efforts to provide contextual learning to trainees. One early NLP effort was the PostDoc and Pindex systems, which could suggest relevant PubMed articles based on curriculum documents [34,35]. One could envision similar approaches applied to clinical notes. Search engines, websites, and PubMed suggest recommendations for similar documents based on document similarity indices, and researchers have demonstrated use of machine learning [36] and topic modeling [37] to cluster documents to each other. Infobuttons have been used to allow users to access information on-demand, incorporate specified logic, and have been shown to reduce time it takes users to find relevant information [38]. Other clinical decision support systems certainly provide significant context-aware, and use of NLP of clinical documents to guide decisions may be nearing implementation [39,40]. In our prior work, we have used NLP in medical education to index medical curricular documents [22] and automatically categorize trainee notes according to competency topics [14,26].

Alternative EHR-based interventions to providing feedback, such as use of the Infobutton standard, have been used to provide just-in-time access to information [31,38,41]. However, our goal was to push such information to all students encountering relevant exposures and to do so outside of the potentially time-pressured environment of clinical care when they might have more time to read or consider the information being presented. One potential criticism of the email approach was that it could constitute "spam" in a student's inbox. While we did not formally survey students, our informal discussions with students after the intervention noted that students for the AD advisor did not find it intrusive. It is interesting to note that no students spontaneously wrote ADs on their patients without first being reminded by the system and that even students that had written ADs did not persist in writing them on subsequent patients without prompting. These observations suggest both that ongoing reminders are necessary but also that education interventions alone may not be sufficient. Greater uptake of ADs could likely be achieved through real-time monitoring of notes, structured forms, or structured/semistructured note formats instead of relying on the students to remember to document these findings [42]. Requiring reminders to improve care is a hallmark of informatics interventions for decades [43].

Our current system recognized the content in the note but did not fully evaluate the quality or competency of students' assessment in these areas. A future system could evaluate for specific physical exam concepts and breadth of differential diagnosis mentioned in notes related to AMS, for example, as a way to evaluate the thoroughness of a student's assessment of the patient. Another approach could be to compare concepts mentioned by students through use of association rule mining constrained by topic (e.g., AMS concepts) and suggest other exam elements, differential diagnoses, etc. that the student could consider in this patient.

5.1. Limitations

Perhaps the biggest limitation of replicating this system is the requirement of available trainee-authored clinical text. However, the rapid adoption and enhancement of EMR systems as part of Meaningful Use requirements [44] should ameliorate the availability of clinical text over time. Even so, the usage of EMRs by medical trainees is heterogeneous. A 2006 survey of 82 internal medicine clerkships found that 48 had explicit policies on student EMR usage: half (23/48) required students to use EMR but the other half (25/48) prohibited allowing students to document in EMR [45]. A 2012 report from the Alliance of Clinical Education found that only two-thirds of institutions allowed students to write notes in the EMR [46]. However, institutions embracing the EMR for educational purposes may find great value to both the trainee and medical team in trainee-authored content in the EMR [47,48].

Other limitations should also be considered. Our review of the AMS advisors was unblinded, so the review may overestimate the true precision of the algorithm. However, comparison of the blinded subset review for AMS documents agreed with the unblinded review for overlapping documents, suggesting the bias may be minimal. The sample size was small, and the recall estimate of the AMS is likely an overestimate. While we have shown that some students receiving AD advisors spontaneously included AD assessment in subsequent notes, we have not demonstrated persistence in knowledge or practice over time. Similarly, we have not shown that the AMS advisor improves knowledge or educational outcomes. We have not surveyed the opinions and attitudes of the students with regard to such a system. It would also be helpful to have formal survey results as to the perceived benefits to the students to augment the formal studies of click-through rates. Our informal focus groups afterward suggested that students did not feel that we were burdening them with too many emails. Finally, this evaluation lacked a formal behavioral model to study student behavior and the impact of the system. The current study focused on changes in competency behaviors but did not formally address learning. Future work for this system and other informatics system addressing medical education outcomes would benefit from such analyses.

6. Conclusions

In summary, we have demonstrated a novel use of NLP to evaluate medical student notes and provide customized feedback, via emails, on their notes. The system was designed within a generalized framework to allow for extension to other topics and requires no interaction from faculty after initial system design. The system achieved good precision and shows promise as an automated method to potentially enhance medical education. Future evaluations should focus on student education outcomes and attitudes toward such systems, as well as expanding interventions.

Conflicts of Interest

The authors have no conflicts of interest.

Acknowledgments

This work was supported in part by grants from the DW Reynolds Foundation, HRSA Geriatric Education Centers Grant #1D31HP08823-01-00, and the National Board of Medical Examiners' Stemmler Foundation. These funding bodies had no role in the execution of the research, evaluation or writing, or final approval of the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2015.06.004.

References

[1] E.C. Corbett, M. Whitcomb, The AAMC Project on the Clinical Education of Medical Students: Clinical Skills Education, 2004. www.im.org/AAIM/PublicPolicy/Docs/ClinicalSKILLs.pdf.
[2] CanMEDS 2000, Extract from the CanMEDS 2000 project societal needs working group report, Med. Teach. 22 (2000) 549–554, http://dx.doi.org/10.1080/01421590050175505.
[3] S.R. Swing, S.G. Clyman, E.S. Holmboe, et al., Advancing resident assessment in graduate medical education, J. Grad. Med. Educ. 1 (2009) 278–286, http://dx.doi.org/10.4300/JGME-D-09-00010.1.
[4] T.J. Nasca, I. Philibert, T. Brigham, et al., The next GME accreditation system—rationale and benefits, N. Engl. J. Med. 366 (2012) 1051–1056, http://dx.doi.org/10.1056/NEJMsr1200117.
[5] Core competencies for the care of older patients: recommendations of the American Geriatrics Society. The Education Committee Writing Group of the American Geriatrics Society, Acad. Med. 75 (2000) 252–255.
[6] B.C. Williams, G. Warshaw, A.R. Fabiny, et al., Medicine in the 21st century: recommended essential geriatrics competencies for internal medicine and family medicine residents, J. Grad. Med. Educ. 2 (2010) 373–383, http://dx.doi.org/10.4300/JGME-D-10-00065.1.
[7] J. La Puma, D. Orentlicher, R.J. Moss, Advance directives on admission clinical implications and analysis of the patient self-determination act of 1990, JAMA 266 (1991) 402–405, http://dx.doi.org/10.1001/jama.1991.03470030102032.
[8] M.E. Whitcomb, Redirecting the assessment of clinical competence, Acad. Med. J. Assoc. Am. Med. Coll. 82 (2007) 527–528, http://dx.doi.org/10.1097/ACM.0b013e31805556f8.
[9] M.H. Davis, R.M. Harden, Competency-based assessment: making it a reality, Med. Teach. 25 (2003) 565–568, http://dx.doi.org/10.1080/0142159032000153842.
[10] G.S. Gotterer, E. Petrusa, S.G. Gabbe, et al., A program to enhance competence in clinical transaction skills, Acad. Med. J. Assoc. Am. Med. Coll. 84 (2009) 838–843, http://dx.doi.org/10.1097/ACM.0b013e3181a81e38.
[11] D.K. Litzelman, A.H. Cottingham, The new formal competency-based curriculum and informal curriculum at Indiana University School of Medicine: overview and five-year analysis, Acad. Med. J. Assoc. Am. Med. Coll. 82 (2007) 410–421, http://dx.doi.org/10.1097/ACM.0b013e31803327f3.
[12] E.F. Dannefer, L.C. Henson, The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine, Acad. Med. J. Assoc. Am. Med. Coll. 82 (2007) 493–502, http://dx.doi.org/10.1097/ACM.0b013e31803ead30.
[13] S.R. Smith, R.H. Dollase, J.A. Boss, Assessing students' performances in a competency-based curriculum, Acad. Med. J. Assoc. Am. Med. Coll. 78 (2003) 97–107.
[14] J.C. Denny, L. Bastarache, E.A. Sastre, et al., Tracking medical students' clinical experiences using natural language processing, J. Biomed. Inform. 42 (2009) 781–789, http://dx.doi.org/10.1016/j.jbi.2009.02.004.
[15] M.I. Langdorf, B.J. Montague, B. Bearie, et al., Quantification of procedures and resuscitations in an emergency medicine residency, J. Emerg. Med. 16 (1998) 121–127.
[16] S.L. Rattner, D.Z. Louis, C. Rabinowitz, et al., Documenting and comparing medical students' clinical experiences, JAMA 286 (2001) 1035–1040, http://dx.doi.org/10.1001/jama.286.9.1035.
[17] T.S. Alderson, N.T. Oswald, Clinical experience of medical students in primary care: use of an electronic log in monitoring experience and in guiding education in the Cambridge Community Based Clinical Course, Med. Educ. 33 (1999) 429–433.
[18] S.B. Bird, R.S. Zarum, F.P. Renzi, Emergency medicine resident patient care documentation using a hand-held computerized device, Acad. Emerg. Med. 8 (2001) 1200–1203.
[19] J. Mattana, M. Charitou, L. Mills, et al., Personal digital assistants: a review of their application in graduate medical education, Am. J. Med. Qual. 20 (2005) 262–267.
[20] P.A. Carney, C.F. Pipas, M.S. Eliassen, et al., An analysis of students' clinical experiences in an integrated primary care clerkship, Acad. Med. 77 (2002) 681–687.
[21] A. Spickard, J. Gigante, G. Stein, et al., Automatic capture of student notes to augment mentor feedback and student performance on patient write-ups, J. Gen. Intern. Med. 23 (2008) 979–984.
[22] J.C. Denny, J.D. Smithers, R.A. Miller, et al., 'Understanding' medical school curriculum content using KnowledgeMap, J. Am. Med. Inf. Assoc. 10 (2003) 351–362.
[23] B.L. Humphreys, D.A. Lindberg, H.M. Schoolman, et al., The unified medical language system: an informatics research collaboration, J. Am. Med. Inf. Assoc. 5 (1998) 1–11.
[24] J.C. Denny, A. Spickard, K.B. Johnson, et al., Evaluation of a method to identify and categorize section headers in clinical documents, J. Am. Med. Inform. Assoc. JAMIA 16 (2009) 806–815, http://dx.doi.org/10.1197/jamia.M3037.
[25] J.C. Denny, R.A. Miller, K.B. Johnson, et al., Development and evaluation of a clinical note section header terminology, AMIA Annu. Symp. Proc. AMIA Symp. (2008) 156–160.
[26] A. Spickard 3rd, H. Ridinger, J. Wrenn, et al., Automatic scoring of medical students' clinical notes to monitor learning in the workplace, Med. Teach. 36 (2014) 68–72, http://dx.doi.org/10.3109/0142159X.2013.849801.

[27] J.C. Denny, P.R. Irani, F.H. Wehbe, et al., The KnowledgeMap project: development of a concept-based medical school curriculum database, AMIA Annu. Symp. Proc. (2003) 195–199.
[28] D.B. Reuben, M. Lee, J.W. Davis, et al., Development and validation of a geriatrics knowledge test for primary care residents, J. Gen. Intern. Med. 12 (1997) 450–452.
[29] M. Lee, L. Wilkerson, D.B. Reuben, et al., Development and validation of a geriatric knowledge test for medical students, J. Am. Geriatr. Soc. 52 (2004) 983–988, http://dx.doi.org/10.1111/j.1532-5415.2004.52269.x.
[30] B.C. Williams, J.T. Fitzgerald, Brief report: brief instrument to assess geriatrics knowledge of surgical and medical subspecialty house officers, J. Gen. Intern. Med. 21 (2006) 490–493, http://dx.doi.org/10.1111/j.1525-1497.2006.00433.x.
[31] S.M. Maviglia, C.S. Yoon, D.W. Bates, et al., KnowledgeLink: impact of context-sensitive information retrieval on clinicians' information needs, J. Am. Med. Inform. Assoc. JAMIA 13 (2006) 67–73, http://dx.doi.org/10.1197/jamia.M1861.
[32] S.T. Rosenbloom, A.J. Geissbuhler, W.D. Dupont, et al., Effect of CPOE user interface design on user-initiated access to educational and patient information during clinical care, J. Am. Med. Inform. Assoc. JAMIA 12 (2005) 458–473, http://dx.doi.org/10.1197/jamia.M1627.
[33] J.C. Denny, R.A. Miller, L.R. Waitman, et al., Identifying QT prolongation from ECG impressions using a general-purpose Natural Language Processor, Int. J. Med. Inf. 78 (Suppl 1) (2009) S34–42.
[34] G.F. Cooper, R.A. Miller, An experiment comparing lexical and statistical methods for extracting MeSH terms from clinical free text, J. Am. Med. Inf. Assoc. 5 (1998) 62–75.
[35] S.L. Kanter, R.A. Miller, M. Tan, et al., Using POSTDOC to recognize biomedical concepts in medical school curricular documents, Bull. Med. Libr. Assoc. 82 (1994) 283–287.
[36] Y. Aphinyanaphongs, I. Tsamardinos, A. Statnikov, et al., Text categorization models for high-quality article retrieval in internal medicine, J. Am. Med. Inform. Assoc. JAMIA 12 (2005) 207–216, http://dx.doi.org/10.1197/jamia.M1641.
[37] Y. Lin, W. Li, K. Chen, et al., A document clustering and ranking system for exploring MEDLINE citations, J. Am. Med. Inform. Assoc. JAMIA 14 (2007) 651–661, http://dx.doi.org/10.1197/jamia.M2215.
[38] G. Del Fiol, P.J. Haug, J.J. Cimino, et al., Effectiveness of topic-specific infobuttons: a randomized controlled trial, J. Am. Med. Inform. Assoc. JAMIA 15 (2008) 752–759, http://dx.doi.org/10.1197/jamia.M2725.
[39] T.D. Imler, J. Morea, T.F. Imperiale, Clinical decision support with natural language processing facilitates determination of colonoscopy surveillance intervals, Clin. Gastroenterol. Hepatol. 12 (2014) 1130–1136, http://dx.doi.org/10.1016/j.cgh.2013.11.025.
[40] K.B. Wagholikar, K.L. MacLaughlin, M.R. Henry, et al., Clinical decision support with automated text processing for cervical cancer screening, J. Am. Med. Inform. Assoc. JAMIA 19 (2012) 833–839, http://dx.doi.org/10.1136/amiajnl-2012-000820.
[41] D.M. Baorto, J.J. Cimino, An 'infobutton' for enabling patients to interpret on-line Pap smear reports, Proc. AMIA Annu. Symp. AMIA Symp. (2000) 47–50.
[42] S.T. Rosenbloom, J.C. Denny, H. Xu, et al., Data from clinical notes: a perspective on the tension between structure and flexible documentation, J. Am. Med. Inform. Assoc. JAMIA 18 (2011) 181–186, http://dx.doi.org/10.1136/jamia.2010.007237.
[43] C.J. McDonald, Protocol-based computer reminders, the quality of care and the non-perfectability of man, N. Engl. J. Med. 295 (1976) 1351–1355, http://dx.doi.org/10.1056/NEJM197612092952405.
[44] D. Blumenthal, M. Tavenner, The 'meaningful use' regulation for electronic health records, N. Engl. J. Med. 363 (2010) 501–504, http://dx.doi.org/10.1056/NEJMp1006114.
[45] M. Mintz, H.J. Narvarte, K.E. O'Brien, et al., Use of electronic medical records by physicians and students in academic internal medicine settings, Acad. Med. 84 (2009) 1698–1704, http://dx.doi.org/10.1097/ACM.0b013e3181bf9d45.
[46] M.M. Hammoud, K. Margo, J.G. Christner, et al., Opportunities and challenges in integrating electronic health records into undergraduate medical education: a national survey of clerkship directors, Teach. Learn. Med. 24 (2012) 219–224, http://dx.doi.org/10.1080/10401334.2012.692267.
[47] M.M. Hammoud, J.L. Dalymple, J.G. Christner, et al., Medical student documentation in electronic health records: a collaborative statement from the Alliance for Clinical Education, Teach. Learn. Med. 24 (2012) 257–266, http://dx.doi.org/10.1080/10401334.2012.692284.
[48] Despite Mounting Role in Medical Student Education, Barriers to EHR Use Remain – AAFP News Now – AAFP. <http://www.aafp.org/news-now/education-professional-development/20130306med-students-ehrs.html> (accessed 10.02.14).